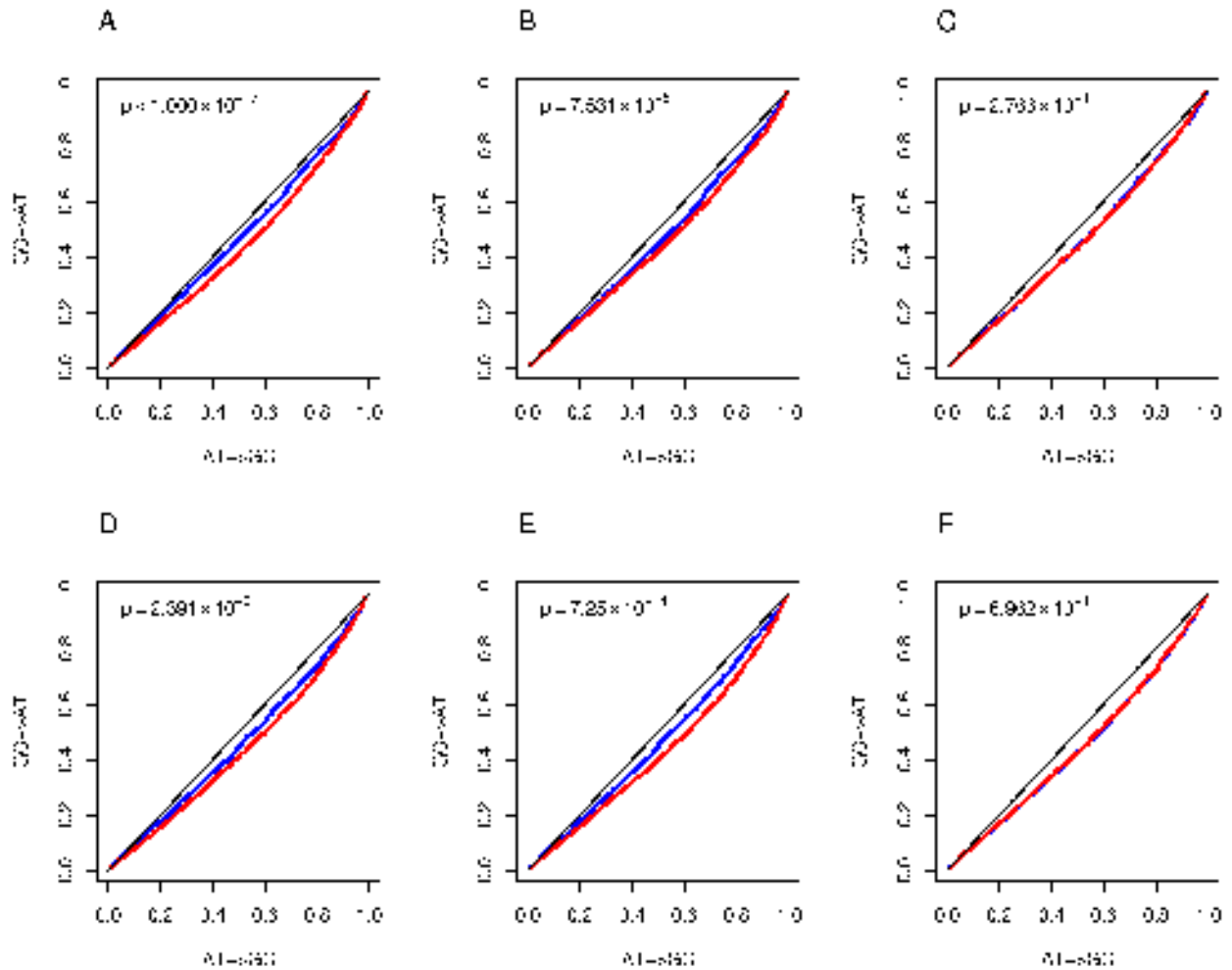


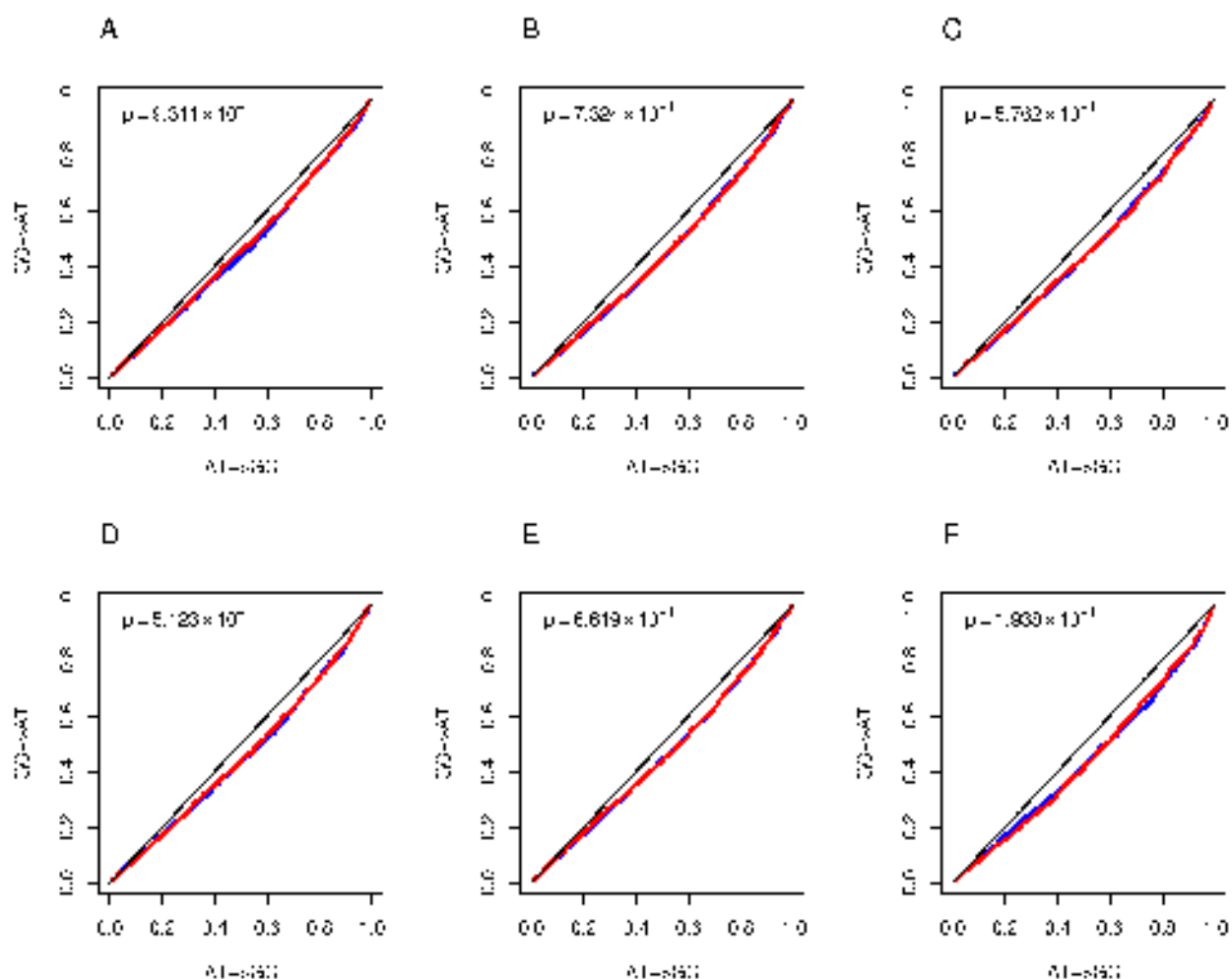
## Additional file 1



Comparison of allele frequency spectra. Results are for 5' (A, B and C) and 3' (D, E and F) intergenic regions. Gene spacers were defined as 5' or 3' intergenic if the definition was consistent for both flanking genes (i.e. in the case of divergent or convergent flanking gene orientations for 5' ad 3' intergenic, respectively).

(A, D) Quantile-quantile plots of GC->AT and AT->GC derived allele frequencies for highly (recombination rate above the 70<sup>th</sup> percentile) and low recombining intergenic regions (recombination rate below the 30<sup>th</sup> percentile) after fixing GC content.

(B, E) The same as (A and D), but in this case we fixed recombination rates and compared high vs low GC regions. (C, F) The same as (A and D), but in this case we fixed both GC content and recombination rates in order to compare regions from highly vs lowly expressed genes.



Comparison of allele frequency spectra. Results are for introns (**B**, **E**) and 5' (**A**, **D**) or 3' (**C**, **F**) intergenic regions. Different measure of gene expression were analysed, namely breadth and peak expression. (**A-C**) Quantile-quantile plots of GC->AT and AT-> GC derived allele frequencies for regions deriving from high (red, peak level higher than the 30<sup>th</sup> percentile) and low (blue, peak level lower than the 30<sup>th</sup> percentiles) expressed genes after fixing GC content and recombination rate. (**D-F**) Quantile-quantile plots of GC->AT and AT-> GC derived allele frequencies for regions deriving from widely (red, breadth  $\geq 68$  tissues) and narrowly (blue, breadth  $\leq 1$  tissues) expressed genes after fixing GC content and recombination rate.